13/prts

DESCRIPTION

Speech Recognition Device, Speech Recognition Method and Speech Recognition Program

Technical Field

The present invention relates to a speech recognition device that recognizes speech issued by a person, a speech recognition method and a speech recognition program.

Background Art

In recent years, there has been significant progress in the technology related to speech recognition. The speech recognition refers to automatic identification of human speech by a computer or a machine. For example, using the speech recognition technique, the computer or machine can be operated in response to human speech or the human speech can be converted into text.

According to a method mainly used in the speech recognition, physical characteristics such as the frequency spectrum of an issued speech are extracted, and compared to pre-stored types of physical characteristics of vowels, consonants, or words. When speech by a number of unspecified speakers is recognized, however, individual differences in the physical characteristics between the speakers impair

accurate speech recognition. If a speech by a particular speaker is recognized, noises caused by changes in the environment such as differences between in the daytime and at night, or changes in the physical characteristics of the speech depending on the health condition of the speaker can lower the speech recognition ratio, in other words accurate speech recognition cannot be performed.

Fig. 13 is a schematic graph showing an example of the relation between the sound level and the recognition ratio in the speech recognition. In the graph shown in Fig. 13, the ordinate represents the recognition ratio (%), while the abscissa represents the sound level (dB). Herein, the sound level means the level of speech power. At 0 dB, for example, the load resistance is 600 Ù, the inter-terminal voltage is 0.775 V, and the power consumption is 1 mW.

As shown in Fig. 13, according to the conventional speech recognition technique, the recognition ratio is lowered when the sound level tends to be lower than -19 dB or higher than -2 dB.

According to the conventional speech recognition technique, the recognition ratio is high in the vicinity of the prestored sound level representing the type of physical characteristics of vowels, consonants, or words. More specifically, the pre-stored sound level and an input sound level are compared for speech recognition, and therefore

equally high recognition ratios do not result for high to low sound levels.

Japanese Utility Model Laid-Open No. 59-60700 discloses a speech recognition device that keeps the input sound level substantially constant using an AGC circuit(Auto Gain Controller) circuit in a micro-amplifier used in inputting sound. Japanese Utility Model Laid-Open No. 01-137497 and Japanese Patent Laid-Open No. 63-014200 disclose a speech recognition device that notifies a speaker of the sound level by some appropriate means, and encourages the speaker to speak in an optimum sound level.

However, by the speech recognition device disclosed by Japanese Utility Model Laid-Open No. 59-60700, unwanted noises other than speech are amplified by the AGC circuit and the amplified noises could lower the recognition ratio. In addition, input speech has accented parts representing the stress of the words on a word-basis. Therefore, if the input sound level is often amplified or not amplified using the AGC circuit, distortions result in the waveform of the speech amplified substantially to a fixed level. The speech waveform distortions distort the accented part of each word representing the stress of the word, which lowers the recognition ratio.

Meanwhile, by the speech recognition devices disclosed by Japanese Utility Model Laid-Open No. 01-137497 and

Japanese Patent Laid-Open No. 63-014200, the sound level input by a speaker might not reach a prescribed value because of changes in the environment or the poor health condition of the speaker. If the speaker speaks in the predetermined sound level, the speech recognition device might not recognize the speech. The level of the speech given by a speaker is for example physical characteristics inherent to the individual, and if the speaker is forced to speak in a different manner, the detected physical characteristic would be different from the original, which could even lower the recognition ratio in the speech recognition.

Disclosure of Invention

It is an object of the present invention to provide a speech recognition device, a speech recognition method and a speech recognition program which can improve the speech recognition ratio regardless of the sound level of a speaker.

A speech recognition device according to one aspect of the present invention includes input means for inputting a digital sound signal, a sound level estimation means for estimating the sound level of a sound period based on the digital sound signal in a part of the sound period input by the input means, sound level adjusting means for adjusting the level of the digital sound signal in the sound period input by the input means based on the sound level estimated by the

sound level estimation means and a preset target level, and speech recognition means for performing speech recognition based on the digital sound signal adjusted by the sound level adjusting means.

In the speech recognition device according to the present invention, a digital sound signal is input by the input means, and the sound level of a sound period is estimated by the sound level estimation means based on the digital sound signal in a prescribed time period of the sound period input by the input means. The level of the digital sound signal in the sound period input by the input means is adjusted based on the sound level estimated by the sound level estimation means and a preset target level, and speech recognition is performed by the speech recognition means based on the digital sound signal adjusted by the sound level adjusting means.

In this case, the sound level of the entire sound period is estimated based on the digital sound signal in a part of the sound period, and the level of the digital sound signal in the sound period is uniformly adjusted based on the estimated sound level and the preset target level. As a result, the accented part of the speech representing the stress of the words uttered by the speaker is not distorted in the speech recognition, which can improve the speech recognition ratio.

The sound level estimation means may estimate the sound level of the sound period based on the digital sound signal in a prescribed time period at the beginning of the sound period input by the input means.

Usually in this case, the sound level of the entire sound period can be determined based on a sound level rising part in a prescribed time period at the beginning of the sound period. Therefore, the sound level is estimated based on the digital sound signal in the prescribed time period at the beginning of the sound period, so that the sound level of the sound period can surely be estimated in a short time period.

The sound level estimation means may estimate the average value of the digital sound signal in a prescribed time period at the beginning of the sound period input by the input means as the sound level of the sound period.

In this case, the sound level of the sound period can more surely be estimated by calculating the average value of the digital sound signal in the prescribed time period at the beginning of the sound period.

The sound level adjusting means may amplify or attenuate the level of the digital sound signal in the sound period input by the input means by an amplification factor determined by the ratio between the preset target level and the sound level estimated by the sound level estimation means.

In this case, the sound level of the sound period can be set to a target level by increasing or attenuating the level of the digital sound signal in the sound period by an amplification factor determined by the ratio between the target level and the estimated sound level.

The speech recognition device may further include a delay circuit that delays the digital sound signal input by the input means so that the digital sound signal input by the input means is applied to the sound level adjusting means together and in synchronization with the sound level estimated by the sound level estimation means.

In this case, the sound level estimation value corresponding to the digital sound signal may be used for adjustment. Thus, the sound level of the sound period can surely be adjusted.

The sound level estimation means may include a sound detector that detects the starting point of sound period input by the input means, a sound level estimator that estimates the sound level of the sound period based on the digital sound signal in a prescribed time period at the beginning of the sound period input by the input means, a hold circuit that holds the sound level estimated by the sound level estimator, and a storing circuit that stores the digital sound signal in the sound period input by the input means in response to the detection by the sound detector and outputs the stored

digital sound signal in the sound period to the sound level adjusting means in synchronization with the sound level held in the hold circuit.

In this case, the starting point of the digital sound signal in the sound period input by the input means is detected by the sound detector, and the sound level of the sound period is estimated by the sound level estimator based on the digital sound signal in the prescribed time period at the beginning of the sound period input by the input means. The sound level estimated by the sound level estimator is held by the hold circuit, the digital sound signal in the sound period input by the input means is stored in the storing circuit in response to the detection of the sound detector, and the stored digital sound signal in the sound period is output to the sound level adjusting means in synchronization with the sound level held in the hold circuit.

In this case, the digital sound signal is stored in the storing circuit from the starting point of the sound period, and the sound level estimation value corresponding to the stored digital sound signal is used for adjusting the sound level. Therefore, the digital sound signal can be adjusted to an accurate sound level and the speech recognition ratio can be improved.

The storing circuit may include first and second buffers that alternately store the digital sound signal in the sound

period input by the input means and alternately output the stored digital sound signal in the sound period to the sound level adjusting means.

In this case, when long speech including a plurality of words is input, the digital sound signal is stored/output alternately to/from the first and second buffers. Thus, the long speech including a plurality of words can be recognized using the first or second buffer having a small capacity.

The speech recognition means may have a result of speech recognition fed back to the sound level adjusting means, and the sound level adjusting means may change the degree of adjusting the sound level based on the result of speech recognition fed back from the speech recognition means.

In this case, an inappropriate sound level adjustment degree may be more optimized by using the result of the speech recognition once again for adjusting the sound level and changing the degree of adjusting the sound level.

The sound level adjusting means may increase the amplification factor for the sound level when speech recognition by the speech recognition means is not possible.

In this case, the sound level not allowing speech recognition can be adjusted to a sound level which allows speech recognition by increasing the amplification factor.

The speech recognition device may further include a non-linear processor that inactivates the sound level

adjusting means when the sound level estimated by the sound level estimation means is within a predetermined range, activates the sound level adjusting means when the sound level estimated by the sound level estimation means is not in the predetermined range, and changes the sound level estimated by the sound level estimation means to a sound level within the predetermined range for application to the sound level adjusting means.

In this case, the sound level can be changed to a sound level within the predetermined range and thus adjusted only when the sound level is not in the predetermined range. Thus, the accented part of the speech representing the stress of the words uttered by the speaker can be prevented from being undesirably distorted.

A speech recognition method according to another aspect of the present invention includes the steps of inputting a digital sound signal, estimating the sound level of a sound period based on the input digital sound signal in a part of the sound period, adjusting the level of the digital sound signal in the sound period based on the estimated sound level and a preset target level, and performing speech recognition based on the adjusted digital sound signal.

In the speech recognition method according to the present invention, a digital sound signal is input, the sound level of a sound period is estimated based on the digital sound

signal in a part of the sound period. The level of the digital sound signal in the sound period is adjusted based on the estimated sound level and a preset target level, and speech recognition is performed based on the adjusted digital sound signal.

In this case, the sound level of the entire sound period is estimated based on the digital sound signal in a part of the sound period, and the level of the digital sound signal in the sound period is uniformly adjusted based on the estimated sound level and a preset target level. As a result, the accented part of the speech representing the stress of the words uttered by the speaker is not distorted in the speech recognition, which can improve the speech recognition ratio.

The step of estimating the sound level may include estimating the sound level of the sound period based on the digital sound signal within a prescribed time period at the beginning of the sound period.

Usually in this case, the sound level of the entire sound period can be determined based on the rising part of the sound level in a prescribed part at the beginning of the sound period. Therefore, The sound level of the sound period can surely be estimated in a short period by estimating the sound level based on the digital sound signal in the prescribed time period at the beginning of the sound period.

The step of estimating the sound level may include

estimating the average value of the digital sound signal in the prescribed time period at the beginning of the sound period as the sound level of the sound period.

In this case, the sound level of the sound period can more surely be estimated by calculating the average value of the digital sound signal in the prescribed time period at the beginning of the sound period.

The step of adjusting the level of the digital sound signal may include amplifying or attenuating the level of the digital sound signal in the sound period by an amplification factor determined by the ratio between the preset target level and the estimated sound level.

In this case, the sound level of the sound period can be set to a target level by increasing or attenuating the level of the digital sound signal in the sound period by an amplification factor determined by the ratio between the target level and the estimated sound level.

The speech recognition method further includes the step of delaying the digital sound signal in the sound period so that the digital sound signal is applied together and in synchronization with the estimated sound level to the step of adjusting the level of the digital sound signal.

In this case, the sound level estimation value corresponding to the digital sound signal may be used for

adjusting the sound level. Thus, the sound level of the sound period can surely be adjusted.

The step of estimating the sound level includes the steps of detecting the starting point of the digital sound signal in the sound period, estimating the sound level of the sound period based on the digital sound signal in a prescribed time period at the beginning of the sound period, holding the estimated sound level, and storing the digital sound signal in the sound period in response to the detection of the starting point of the digital sound signal and outputting the stored digital sound signal in the sound period in synchronization with the held sound level.

In this case, the starting point of the digital sound signal in the sound period is detected, and the sound level of the sound period is estimated based on the digital sound signal in a prescribed time period at the beginning of the sound period. The estimated sound level is held, the digital sound signal in the sound period is stored in response to the detection of the starting point of the digital sound signal in the sound period and the stored digital sound signal in the sound period is output in synchronization with the held sound level.

In this case, the digital sound signal is stored in the storing circuit from the starting point of the sound period, and the sound level is adjusted using the sound level

estimation value corresponding to the stored digital sound signal. Thus, the sound level can be adjusted to an accurate sound level, which can improve the speech recognition ratio.

The storing step includes the step of storing the digital sound signal in the sound period alternately to first and second buffers and outputting the stored digital sound signal in the sound period alternately from the first and second buffers.

In this case, when long speech including a plurality of words is input, the digital sound signal is stored/output alternately to/from the first and second buffers. Thus, the long speech including a plurality of words can be recognized using the first or second buffer having a small capacity.

The step of performing the speech recognition may include the step of feeding back a result of speech recognition during the step of adjusting the level of the digital sound signal, and the step of adjusting the level of the digital sound signal may include changing the degree of adjusting the sound level based on the fed back result of speech recognition.

In this case, only an inappropriate sound level adjustment degree may be more optimized by using the result of the speech recognition once again for adjusting the sound level and changing the degree of adjusting sound level.

The step of adjusting the level of the digital sound

signal may include increasing the amplification factor for the sound level when the speech recognition is not possible.

In this case, the sound level not allowing speech recognition can be adjusted to a sound level which allows speech recognition by increasing the amplification factor for the sound level.

The speech recognition method further includes the step of inactivating the step of adjusting the level of the digital sound signal when the estimated sound level is within a predetermined range, while activating the adjusting step when the estimated sound level is not in the predetermined range, and changing the estimated sound level to a sound level within the predetermined range for use in adjusting the level of the digital sound signal.

In this case, the sound level can be changed to a sound level within the predetermined range and thus adjusted only when the sound level is not in the predetermined range. Thus, the accented part of the speech representing the stress of the words uttered by the speaker can be prevented from being undesirably distorted.

A speech recognition program according to another aspect of the present invention enables a computer to execute the steps of inputting a digital sound signal, estimating the sound level of the sound period based on the input digital sound signal in a part of the sound period, adjusting the level

of the input digital sound signal in the sound period based on the estimated sound level and a preset target level, and performing speech recognition based on the adjusted digital sound signal.

In the speech recognition program according to the present invention, the digital sound signal is input and the sound level of a sound period is estimated based on the input digital sound signal in a predetermined time period of the sound period. The level of the input digital sound signal in the sound period is adjusted based on the estimated sound level and a preset target value, and speech recognition is performed based on the adjusted digital sound signal.

In this case, the sound level of the entire sound period is estimated based on the digital sound signal in a part of the sound period, and the level of the digital sound signal in the sound period is uniformly adjusted based on the estimated sound level and the preset target level. As a result, the accented part of the speech representing the stress of the words uttered by the speaker is not distorted in the speech recognition. This can increase the speech recognition ratio.

According to the present invention, the sound level of the entire sound period is estimated based on the digital sound signal in a part of the sound period, and the level of the digital sound signal in the sound period is uniformly

adjusted based on the estimated sound level and a preset target level. As a result, the accented part of the speech representing the stress of the words uttered by the speaker is not distorted in the speech recognition. This can increase the speech recognition ratio.

Brief Description of the Invention

Fig. 1 is a block diagram of a speech recognition device according to one embodiment of the present invention;

Fig. 2 is a block diagram of the configuration of a computer to execute a speech recognition program;

Fig. 3 is a waveform chart showing the speech spectrum of a word "*ragubi*" uttered by a speaker;

Fig. 4 is a block diagram of a speech recognition device according to a second embodiment of the present invention;

Fig. 5(a) is a waveform chart for the output of a microphone in Fig. 4, while Fig. 5(b) is a graph showing the ratio of the sound signal (signal component) to noise component;

Fig. 6 is a flowchart showing the operation of a sound detector shown in Fig. 4;

Fig. 7 is a schematic diagram showing input/output of a digital sound signal to/from buffers when a speaker utters two words;

Fig. 8 is a block diagram showing an example of a speech recognition device according to a third embodiment of the present invention;

Fig. 9 is a flowchart for use in illustration of the operation of the sound level adjusting feedback unit shown in Fig. 8 when the sound level is adjusted;

Fig. 10 is a block diagram showing an example of a speech recognition device according to a fourth embodiment of the present invention;

Fig. 11 is a graph for use in illustration of the relation between a sound level estimation value input to a signal non-linear processor and the recognition ratio in the speech recognition unit in Fig. 10;

Fig. 12 is a flowchart for use in illustration of the processing operation of the signal non-linear processor; and

Fig. 13 is a schematic graph showing an example of the relation between the sound level and the recognition ratio in the speech recognition.

Best Mode for Carrying Out the Invention

First Embodiment

Fig. 1 is a block diagram of an example of a speech recognition device according to one embodiment of the present invention.

As shown in Fig. 1, the speech recognition device

includes a microphone 1, an A/D (analog-digital) converter 2, a signal delay unit 3, a sound level estimator 4, a sound level adjuster 5 and a speech recognition unit 6.

As shown in Fig. 1, speech issued by a speaker is collected by the microphone 1. The collected speech is converted into an analog sound signal SA by the function of the microphone 1 for output to the A/D converter 2. The A/D converter 2 converts the applied analog signal SA into a digital sound signal DS for output to the signal delay unit 3 and the sound level estimator 4. The sound level estimator 4 calculates a sound level estimation value LVL based on the applied digital sound signal DS. Herein, the sound level refers to the level of sound power (sound energy). How to calculate the sound level estimation value LVL will later be described.

The signal delay unit 3 applies the digital sound signal DS delayed by a period corresponding to a prescribed sound level rising time TL which will be described to the sound level adjuster 5. The sound level adjuster 5 adjusts the sound level of the digital sound signal DS applied from the signal delay unit 3 in synchronization with the sound level estimation value LVL applied from the sound level estimator 4. The sound level adjuster 5 applies an output CTRL_OUT after the adjustment of the sound level to the speech recognition unit 6. The speech recognition unit 6 performs speech recognition

based on the output CTRL_OUT after the adjustment of the sound level applied from the sound level adjuster 5.

In the speech recognition device according to the first embodiment, the microphone 1 and the A/D (analog-digital) converter 2 correspond to the input means, the signal delay unit 3 to the delay circuit, the sound level estimator 4 to the sound level estimation means, the sound level adjuster 5 to the sound level adjusting means, and the speech recognition unit 6 to the speech recognition means.

Note that the signal delay unit 3, the sound level estimator 4, the sound level adjuster 5 and the speech recognition unit 6 may be implemented by the signal delay circuit, the sound level estimation circuit, the sound level adjusting circuit and the speech recognition circuit, respectively. Meanwhile, the signal delay unit 3, the sound level estimator 4, the sound level adjuster 5 and the speech recognition unit 6 may be implemented by a computer and a speech recognition program.

Such a computer to execute the speech recognition program will now be described. Fig. 2 is a block diagram of the configuration of the computer to execute the speech recognition program.

The computer includes a CPU (Central Processing Unit) 500, an input/output device 501, a ROM (Read Only Memory) 502, a RAM (Random Access Memory) 503, a recording medium 504, a

recording medium drive 505, and an external storage 506.

The input/output device 501 transmits/receives information to/from other devices. The digital sound signal DS from the A/D converter 2 in Fig. 1 is input to the input/output device 501 according to the embodiment. The ROM 502 is recorded with system programs. The recording medium drive 505 is of a CD-ROM drive, a floppy disc drive, or the like and reads/writes data from/to a recording medium 504 such as a CD-ROM and a floppy disc. The recording medium 504 is recorded with speech recognition programs. The external storage 506 is of a hard disc and the like and is recorded with a speech recognition program read from the recording medium 504 through the recording medium drive 505. The CPU 500 executes the speech recognition program stored in the external storage 506 on the RAM 503. Thus, the functions of the signal delay unit 3, the sound level estimator 4, the sound level adjuster 5 and the speech recognition unit 6 in Fig. 1 are executed.

Now, a method of calculating the sound level estimation value LVL by the sound level estimator 4 in Fig. 1 and a method of adjusting the sound level by the sound level adjuster 5 will be described.

The method of calculating the sound level estimation value LVL by the sound level estimator 4 will be described first. The digital sound signal DS input to the sound level

estimator 4 is represented as DS(x)(x=1, 2, ..., Q) where x indicates Q time points in the rising time TL for a predetermined sound level, and DS(x) indicates the value of the digital sound signal DS at the Q time points. In this case, the sound level estimation value LVL is expressed as follows:

$$LVL = (\Sigma |DS(x)|)/Q \quad \ldots(1)$$

In the expression (1), the sound level estimation value LVL is the average value produced by dividing the cumulative sum of the absolute values of the digital sound signal DS (x) at the Q time points in the rising time TL of the predetermined sound level by Q. Thus, the sound level estimation value LVL is calculated in the sound level estimator 4.

Now, the method of adjusting the sound level by the sound level adjuster 5 will now be described. In the sound level adjuster 5, a target value for a predetermined sound level is indicated as TRG_LVL. In this case, the adjusted value for the sound level LVL_CTRL is expressed as follows:

$$LVL\_CTRL = TGR\_LVL/LVL \quad \ldots(2)$$

In the expression (2), the adjusted value LVL_CTRL for the sound level is calculated by dividing the target value

TRG_LVL for the predetermined sound level by the sound level estimation value LVL.

The output CTRL_OUT after the adjustment of the sound level is expressed using the adjusted value LVL_CTRL for the sound level as follows:

$$CTRL\_OUT(X)=DS(X)\times LVL\_CTRL \quad \ldots(3)$$

where X represents time. In the expression (3), the output CTRL_OUT(X) after the adjustment of the sound level is produced by multiplying the digital sound signal DS(X) at a predetermined sound level rising time TL by the adjusted value LVL_CTRL for the sound level. Thus, the sound level adjuster 5 adjusts the sound level and applies the resulting output CTRL_OUT (X) to the speech recognition unit 6.

The predetermined rising time TL for the sound level in the signal delay unit 3 shown in Fig. 1 will now be described in conjunction with the drawings.

Fig. 3 is a waveform chart showing the speech spectrum of a word "*ragubi*" uttered by a speaker. In Fig. 3, the ordinate represents the sound level, while the abscissa represents time.

As shown in Fig. 3, in the speech spectrum of the word "*ragubi*," the sound level of the "*ra*" part is high. More specifically, the high point in the sound level corresponds

to the part where the accent representing the stress of each word lies. Here, as shown in Fig. 3, the time from the starting point TS when a word is uttered by the speaker to the time point when the peak value P of the sound level is reached is the sound level rising time TL. In general, the sound level rising time TL is in the range from 0 sec to 100 msec, and the sound level rising time TL according to the embodiment of the invention is for example 100 msec.

If for example the sound level rising time TL is set to a shorter period, the speech recognition ratio is lowered. As shown in Fig. 3, assume that the speaker utters the word "*ragubi*," and a shorter sound level rising time denoted by TL' is set. In this case, simply delaying the digital sound signal DS input to the signal delay unit 3 shown in Fig. 1 by the rising time TL' does not allow an appropriate sound level estimation value LVL to be calculated by the sound level estimator 4. A sound level estimation value lower than the intended target sound level estimation value LVL is produced. Then, the sound level estimation value lower than the target value is provided to the sound level adjuster 5, and the sound level value of the digital sound signal DS is adjusted incorrectly by the sound level adjuster 5. Thus, the incorrect digital sound signal DS is input to the speech recognition unit 6, which lowers the speech recognition ratio.

As described above, the sound level rising time TL at the beginning of a sound period is set to 100 msec at the signal delay unit 3, so that the sound level of the entire sound period can be calculated by the sound level estimator 4. Thus, the level of the digital sound signal DS of the sound period is uniformly adjusted. As a result, the accented part of the speech representing the stress of the words uttered by the speaker is not distorted in the speech recognition, which increases the speech recognition ratio.

Second Embodiment

A speech recognition device according to a second embodiment of the invention will now be described in conjunction with the accompanying drawings.

Fig. 4 is a block diagram of a speech recognition device according to the second embodiment of the present invention.

As shown in Fig. 4, the speech recognition device includes a microphone 1, an A/D converter 2, a sound level estimator 4, a sound level adjuster 5, a speech recognition unit 6, a sound detector 7, a sound level holder 8, selectors 11 and 12, and buffers 21 and 22.

As shown in Fig. 4, speech issued by a speaker is collected by the microphone 1. The collected speech is converted into an analog sound signal SA by the function of the microphone 1 for output to the A/D converter 2. The A/D converter 2 converts the applied analog sound signal SA into

a digital sound signal DS for application to the sound level estimator 4, the sound detector 7, and the selector 11. The sound level estimator 4 calculates the sound level estimation value LVL based on the applied digital sound signal DS. The method of calculating the sound level estimation value LVL by the sound level estimator 4 according to the second embodiment is the same as the method of calculating the sound level estimation value LVL by the sound level estimator 4 according to the first embodiment.

The sound level estimator 4 calculates a sound level estimation value LVL for each word based on the digital sound signal DS applied from the A/D converter 2, and sequentially applies the resulting sound level estimation value LVL to the sound level holder 8. Here, the sound level holder 8 holds the previous sound level estimation value LVL in a holding register provided in the sound holder 8 until the next sound level estimation value LVL calculated by the sound level estimator 4 is applied and overwrites each new sound level estimation value LVL applied from the sound level estimator 4 in the holding register holding the previous sound level estimation value LVL. The holding register has a data capacity M.

Meanwhile, the sound detector 7 detects the starting point TS of the sound in Fig. 3 based on the digital sound signal DS applied from the A/D converter 2, and applies a

control signal CIS1 to the selector 11 so that the digital sound signal DS is applied to the buffer 21, and a control signal CB1 to the buffer 21 so that the digital sound signal DS applied from the selector 11 is stored therein. The buffers 21 and 22 both have a capacity L.

The selector 11 applies the digital sound signal DS applied from the A/D converter 2 to the buffer 21 in response to the control signal CIS1 applied from the sound detector 7. The buffer 21 stores the digital sound signal DS applied through the selector 11 in response to the control signal CB1 applied from the sound detector 7. The buffer 21 applies a full signal F1 to the sound detector 7 when it has stored the digital sound signal DS as much as the storable capacity L. Thus, the sound detector 7 applies a control signal SL1 to cause the sound level holder 8 to output the sound level estimation value LVL through the buffer 21.

The sound detector 7 applies a control signal CIS2 to the selector 11 in response to the full signal F1 applied from the buffer 21 so that the digital sound signal DS applied from the A/D converter 2 is applied to the buffer 22 and a control signal CB2 to the buffer 22 so that the digital sound signal DS applied from the selector 11 is stored therein. In addition, the sound detector 7 applies a control signal CBO1 to the buffer 21 and a control signal COS1 to the selector 12.

The selector 11 applies the digital sound signal DS applied from the A/D converter 2 to the buffer 22 in response to the control signal CIS2 applied from the sound detector 7. The buffer 22 stores the digital sound signal DS applied through the selector 11 in response to the control signal CB2 applied from the sound detector 7.

Meanwhile, the buffer 21 applies the digital sound signal DS stored in the buffer 21 to the sound level adjuster 5 through the selector 12 in response to the control signal CBO1 applied from the sound detector 7.

The buffer 22 stores the digital sound signal DS applied through the selector 11 in response to the control signal CB2 applied from the sound detector 7. The buffer 22 applies the full signal F2 to the sound detector 7 when it has stored the digital sound signal DS as much as its storable capacity L. Thus, the sound detector 7 applies a control signal SL2 through the buffer 22 to cause the sound level holder 8 to output the sound level estimation value LVL.

The sound detector 7 applies the control signal CIS1 to the selector 11 in response to the full signal F2 applied from the buffer 22 so that the digital sound signal DS applied from the A/D converter 2 is applied to the buffer 21. The sound detector 7 applies a control signal CBO2 to the buffer 22 and a control signal COS2 to the selector 12.

Meanwhile, the buffer 22 applies the digital sound

signal DS stored in the buffer 22 to the sound level adjuster 5 through the selector 12 in response to the control signal CBO2 applied from the sound detector 7.

The sound level holder 8 applies the sound level estimation value LVL held by the holding register inside to the sound level adjuster 5 in response to the control signal SL1 applied from the buffer 21 or the control signal SL2 applied from the buffer 22. Here, the capacity M of the holding register provided in the sound level holder 8 and the capacity L of the buffers 21 and 22 are substantially the same, and therefore the sound level estimation value LVL corresponding to the digital sound signal DS applied through the selector 12 is output from the sound level holder 8.

The sound level adjuster 5 adjusts the digital sound signal DS obtained through the selector 12 based on the sound level estimation value LVL applied from the sound level holder 8. The method of adjusting the digital sound signal DS by the sound level adjuster 5 according to the second embodiment is the same as the method of adjusting the digital sound signal DS by the sound level adjuster 5 according to the first embodiment. The sound level adjuster 5 applies the sound level adjusted output CTRL_OUT to the speech recognition unit 6. The speech recognition unit 6 performs speech recognition based on the sound level adjusted output CTRL_OUT applied from the sound level adjuster 5.

In the speech recognition device according to the second embodiment, the microphone 1 and the A/D (analog-digital) converter 2 correspond to the input means, the sound level estimator 4 to the sound level estimation means, the sound level adjuster 5 to the sound level adjusting means, the speech recognition unit 6 to the speech recognition means, the speech detector 7 to the sound detector, the sound level holder 8 to the hold circuit, and the buffers 21 and 22 to the storing circuit.

Fig. 5(a) is a waveform chart for the output of the microphone 1 in Fig. 4, while Fig. 5(b) is a graph showing the ratio of the sound signal (signal component) S to noise component N (S/N).

As shown in Fig. 5(a), the output waveform of the microphone 1 consists of the noise component and the sound signal. The sound period including the sound signal has a high sound level value in the output waveform.

As shown in Fig. 5(b), the sound detector 7 in Fig. 4 determines any period having a low S/N ratio, the ratio of the sound signal (speech component) to the noise component as a noise period, while the detector determines any period having a high S/N ratio as a sound period.

Fig. 6 is a flowchart showing the operation of the sound detector 7 shown in Fig. 4.

As shown in Fig. 6, the sound detector 7 determines

whether or not the input digital sound signal DS is a sound signal (step S61). If the input digital sound signal DS is not a sound signal, the sound detector 7 stands by until the following digital sound signal DS input is determined as a sound signal. Meanwhile, if the input digital sound signal DS is determined as a sound signal, the sound detector 7 applies the control signal CIS1 to the selector 11 in Fig. 4 so that the digital sound signal DS applied to the selector 11 is applied to the buffer 21 (step S62). The sound detector 7 applies the control signal CB1 to the buffer 21 so that the digital sound signal DS is stored in the buffer 21 (step S63).

The sound detector 7 then determines whether or not the full signal F1 which is output when the digital sound signal DS as much as the storable capacity L by the buffer 21 has been stored is received (step S64). The sound detector 7 repeats the step S63 before the full signal F1 is not received from the buffer 21. Meanwhile, the sound detector 7 applies the control signal CIS2 to the selector 11 in Fig. 4 in response to the full signal F1 received from the buffer 21 so that the digital sound signal DS applied to the selector 11 is applied to the buffer 22 (step S65). The sound detector 7 applies the control signal CB2 to the buffer 22 so that the buffer 22 stores the digital sound signal DS (step S66). The sound detector 7 outputs the control signals CIS2 and CB2, and then applies the control signal COS1 to the selector 12

so that the stored digital sound signal DS applied from the buffer 21 is applied to the sound level adjuster 5 (step S67).

The sound detector 7 then applies the control signal SL1 to the sound level holder 8 through the buffer 21 (step S68). The sound level holder 8 applies to the sound level adjuster 5 the sound level estimation value LVL repeatedly stored in the holding register in the sound level holder 8 in response to the control signal SL1 applied through buffer 21.

Then, the sound detector 7 applies the control signal CBO1 to the buffer 21, so that the stored digital sound signal DS is output to the sound level adjuster 5 (step S69). The sound detector 7 then determines whether or not the digital sound signal DS stored in the buffer 21 is entirely output to the sound level adjuster 5 (step S70). Here, if the digital sound signal DS is not entirely output from the buffer 21, the control signal CBO1 is once again applied to the buffer 21, so that the stored digital sound signal DS is output to the sound level adjuster 5. Meanwhile, when the digital sound signal DS stored in the buffer 21 is entirely output, the sound detector 7 applies a control signal CR to the buffer 21 so that the data in the buffer is erased (cleared) (step S71).

Fig. 7 is a schematic chart showing input/output of the digital sound signal DS to/from the buffers 21 and 22 when a speaker utters two words.

As shown in Fig. 7, the buffer 21 is provided with the

control signal CB1 from the sound detector 7 at the beginning of one word W1 in a sound period S, so that the digital sound signal DS starts to be input to the buffer 21. Herein, the buffers 21 and 22 are FIFO (First In First Out) type memories, and have substantially the same memory capacity L.

The digital sound signal DS is input to the buffer 21 for almost the entire one word W1, and once the digital sound signal DS as much as the capacity L storable in the buffer 21 has been stored, the buffer 21 outputs the full signal F1 to the sound detector 7. The buffer 21 outputs the full signal F1 and then outputs the digital sound signal DS stored in buffer 21 in response to the control signal CBO1 applied from the sound detector 7. Meanwhile, the buffer 22 starts to store the digital sound signal DS in response to the control signal CB2 applied from the sound detector 7.

The buffer 22 outputs the full signal F2 to the sound detector 7 when the digital sound signal DS as much as its storable capacity L has been stored. Meanwhile, the digital sound signal DS stored in the buffer 21 during the storing of the signal in the buffer 22 is entirely output to the sound level adjuster 5 and then the data in the buffer 21 is all erased (cleared) in response to the control signal CR applied from the sound detector 7. Thus, the control signal CB1 to cause the digital sound signal DS to be once again stored is applied to the buffer 21 from the sound detector 7.

As described above, the digital sound signal is stored from the starting point of a sound period, and a sound level estimation value corresponding to the stored digital sound signal may be used to accurately adjust the sound level. As a result, the speech recognition can be adjusted based on the accurate sound level, so that the speech recognition ratio can be improved.

If a digital sound signal DS for a long period including a plurality of words is input, storing and output operations can alternatively be performed. In this way, the speech recognition can be performed using a buffer having only a small capacity.

Note that while the buffers are used according to the embodiment of the invention, storing circuits of other kinds may be used. Furthermore, the buffer may be provided with a counter inside, and the counter in the buffer may be monitored by the sound detector 7, and the full signal F1 or F2 or the control signal CR may be output.

Third Embodiment

Fig. 8 is a block diagram showing an example of a speech recognition device according to a third embodiment of the present invention.

As shown in Fig. 8, the speech recognition device includes a microphone 1, an A/D (analog-digital) converter 2, a signal delay unit 3, a sound level estimator 4, a sound

level adjusting feedback unit 9, and a speech recognition feedback unit 10.

As shown in Fig. 8, speech issued by a speaker is collected by the microphone 1. The collected speech is converted into an analog sound signal SA by the function of the microphone 1 for output to the A/D converter 2. The A/D converter 2 converts the analog sound signal SA into a digital sound signal DS for application to the signal delay unit 3 and the sound level estimator 4. The sound level estimator 4 calculates a sound level estimation value LVL based on the applied digital sound signal DS. Here, the method of calculating the sound level estimation value LVL by the sound level estimator 4 according to the third embodiment is the same as the method of calculating the sound level estimation value LVL by the sound level estimator 4 according to the first embodiment.

The sound level estimator 4 calculates the sound level estimation value LVL for application to the sound level adjusting feedback unit 9. The sound level adjusting feedback unit 9 adjusts the level of the digital sound signal DS applied from the signal delay unit 3 based on and in synchronization with the sound level estimation value LVL applied from the sound level estimator 4. The sound level adjusting feedback unit 9 applies to the speech recognition feedback unit 10 an output CTRL_OUT after the adjustment of the sound level. The

speech recognition feedback unit 10 performs speech recognition based on the adjusted output CTRL_OUT applied from the sound level adjusting feedback unit 9, and applies the sound level control signal RC to the sound level adjusting feedback unit 9 when the speech recognition is not successful. The operation of the sound level adjusting feedback unit 9 and speech recognition feedback unit 10 will be described later.

In the speech recognition device according to the third embodiment, the microphone 1 and the A/D (analog-digital) converter 2 correspond to the input means, the signal delay unit 3 to the delay circuit, the sound level estimator 4 to the sound level estimation means, the sound level adjusting feedback unit 9 to the sound level adjusting means, and the speech recognition feedback unit 10 to the speech recognition means.

Fig. 9 is a flowchart for use in illustration of the operation of the sound level adjusting feedback unit 9 shown in Fig. 8 when the sound level is adjusted.

As shown in Fig. 9, the sound level adjusting feedback unit 9 determines whether or not the sound level control signal RC by the speech recognition feedback unit 10 is input (step S91). If the sound level control signal RC is not input by the speech recognition feedback unit 10, the sound level adjusting feedback unit 9 stands by until it is determined

that the sound level control signal RC is input from the speech recognition feedback unit 10. Meanwhile, if it is determined that the sound level control signal RC is input from the speech recognition feedback unit 10, the sound level adjusting feedback unit 9 adds 1 to the variable K (step S92).

Here, sound level target values in a plurality of levels are preset, and the variable K represents the number of the levels. According to the third embodiment, the variable K has a value in the range from 1 to R, and the sound level target value TRG_LVL(K) can be TRG_LVL(1), TRG_LVL(2),..., or TRG_LVL(R).

The sound level adjusting feedback unit 9 then determines whether or not the variable K is larger than the maximum value R (step S93). Here, the sound level adjusting feedback unit 9 determines that the variable K is larger than the maximum value R, the sound level adjusting feedback unit 9 returns the variable K to the minimum value 1 (step S94), and sets the sound level target value TRG_LVL to TRG_LVL(1) (step S95).

Meanwhile, if the sound level adjusting feedback unit 9 determines that the variable K is the maximum value R or less, the sound level adjusting feedback unit 9 sets the sound level target value TRG_LVL to TRG_LVL(K)(step S95).

Assume that the sound level target value TRG_LVL is initially set for example to TRG_LVL(2). If then the speech

recognition feedback unit 10 has failed to recognize speech or speech recognition is unsuccessful, the control signal RC is output to the sound level adjusting feedback unit 9. The sound level adjusting feedback unit 9 changes the sound level target value TRG_LVL(2) to the sound level target value TRG_LVL(3), and waits for speech input again from the speaker.

In this way, the sound level target value TRG_LVL is sequentially changed to the sound level target value TRG_LVL(2), TRG_LVL(3) and TRG_LVL(4), and when the speech recognition is successfully performed, the sound level target value TRG_LVL at the time is fixed. If the sound level target value TRG_LVL is set to the maximum value TRG_LVL(R), and still the speech recognition is not successful, the sound level target value TRG_LVL is returned to the minimum value TRG_LVL(1), and speech input again from the speaker is waited.

Thus, the sound level target value TRG_LVL is set to the optimum value for speech recognition.

As described above, when the speech recognition is not successfully performed, the degree of the sound level adjustment can sequentially be raised again by the sound level adjusting feedback unit 9. If the sound level is adjusted to the degree of the predetermined maximum sound level value, the sound level can be returned to the minimum level and once again the degree of adjustment can sequentially be raised. Thus, when the speech recognition is not successful because

the degree of sound level adjustment is not appropriate, the degree can repeatedly and sequentially be changed, so that the speech recognition ratio can be improved.

Note that according to the above described embodiment, after unsuccessful speech recognition, the target value TRG_LVL(K) for the sound level is sequentially changed based on speech input again from the speaker. Meanwhile, the invention is not limited to this, and means for holding speech input may be provided and upon unsuccessful speech recognition, the speech input held by the speech input holding means may be used to sequentially change the sound level target TRG_LVL(K).

Fourth Embodiment

Fig. 10 is a block diagram showing an example of a speech recognition device according to a fourth embodiment of the present invention.

As shown in Fig. 10, the speech recognition device includes a microphone 1, an A/D(analog-digital) converter 2, a signal delay unit 3, a sound level estimator 4, a sound level adjuster 5, a speech recognition unit 6 and a signal non-linear processor 11.

As shown in Fig. 10, speech issued by a speaker is collected by the microphone 1. The collected speech is converted into an analog sound signal SA by the function of microphone 1 for output to the A/D converter 2. The A/D

converter 2 converts the analog sound signal SA into a digital sound signal DS for application to the signal delay unit 3 and the sound level estimator 4. The sound level estimator 4 calculates a sound level estimation value LVL based on the applied digital sound signal DS. Here, the method of calculating the sound level estimation value LVL by the sound level estimator 4 according to the fourth embodiment is the same as the method of calculating the sound level estimation value LVL by the sound level estimator 4 according to the first embodiment. The sound level estimator 4 applies the digital sound signal DS and the sound level estimation value LVL to the signal non-linear processor 11. The signal non-linear processor 11 performs non-linear processing as will be described based on the sound level estimation value LVL applied from the sound level estimator 4, and applies the sound level estimation value LVL after the non-linear processing to the sound level adjuster 5.

Meanwhile, the signal delay unit 3 applies the digital sound signal DS delayed by a period corresponding to the sound level rising time TL to the sound level adjuster 5. Here, the delay corresponding to the sound level rising time TL according to the fourth embodiment is 100 msec. The sound level adjuster 5 performs the sound level adjustment of the digital sound signal DS applied from the signal delay unit 3 based on the sound level estimation value LVL applied from

the signal non-linear processor 11. The sound level adjuster 5 applies the sound level adjusted output CTRL_OUT to the speech recognition unit 6. The speech recognition unit 6 performs speech recognition based on the sound level adjusted output CTRL_OUT applied from the sound level adjuster 5.

In the speech recognition device according to the fourth embodiment, the microphone 1 and the A/D (analog-digital) converter 2 correspond to the input means, the signal delay unit 3 to the delay circuit, the sound level estimator 4 to the sound level estimation means, the sound level adjuster 5 to the sound level adjusting means, the speech recognition unit 6 to the speech recognition means, and the signal non-linear processor 11 to the non-linear processor.

Fig. 11 is a graph for use in illustration of the relation between the sound level estimation value LVL input to the signal non-linear processor 11 in Fig. 10 and the recognition ratio in the speech recognition unit 6 in Fig. 10.

As shown in Fig. 11, the recognition ratio in the speech recognition unit 6 in Fig. 10 depends on the sound level estimation value LVL. When the sound level estimation value LVL is in the range from -19 dB to -2 dB, the recognition ratio is 80% or more. When the sound level estimation value LVL is particularly low (at most -19 dB) or high (at least -2 dB), the speech recognition ratio abruptly drops.

Consequently, in the signal non-linear processor 11 according to the fourth embodiment of the invention, the input sound level estimation value LVL is adjusted to be in the range from -19 dB to -2 dB.

Fig. 12 is a flowchart for use in illustration of the processing operation of the signal non-linear processor 11.

As shown in Fig. 12, the signal non-linear processor 11 determines whether or not the sound level estimation value LVL input from the sound level estimator 4 is in the range from -19 dB to -2 dB (step S101).

When the signal non-linear processor 11 determines that the input sound level estimation value LVL is from -19 dB to -2 dB, the sound level adjuster 5 is inactivated. More specifically, in the sound level adjuster 5, the sound level adjusting value LVL_CTRL is 1 in the expression (2) in this case.

Meanwhile, when the signal non-linear processor 11 determines that the input sound level estimation value LVL is not in the range from -19 dB to -2 dB, the sound level estimation value LVL is set to -10 dB (step S102).

As described, the signal non-linear processor 11 sets the sound level estimation value LVL to allow the recognition ratio to be at least 80%, and therefore the recognition ratio of the input digital sound signal DS in the speech recognition unit 6 can be improved. More specifically, only when the sound

level estimation value LVL is not in the predetermined range, the sound level estimation value is changed to a sound level estimation value within the predetermined range for adjusting the sound level. Meanwhile, when the sound level estimation value is within the predetermined range, the amplification factor is set to 1 in the sound level adjuster 5 to inactivate the sound level adjuster 5, so that the sound level is not adjusted. Thus, speech recognition can readily be performed without undesirably distorting the accented part of the speech representing the stress of the words uttered by the speaker, so that the recognition ratio can be improved.

Note that in the above embodiment, the sound level estimation value is adjusted within the range from -19 dB to -2 dB, while the invention is not limited to this, and the value may be adjusted to a preset sound level estimation value in the speech recognition or a sound level estimation value which allows a higher recognition ratio.